PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G10L 7/06, 9/18	A1	(11) International Publication Number: WO 00/13172 (43) International Publication Date: 9 March 2000 (09.03.00)

(21) International Application Number: PCT/NZ99/00143

(22) International Filing Date: 27 August 1999 (27.08.99)

(30) Priority Data:
331639 28 August 1998 (28.08.98) NZ

(71) Applicant (*for all designated States except US*): SIGMA AUDIO RESEARCH LIMITED [NZ/NZ]; 13 Vienna Place, Birkenhead, Auckland (NZ).

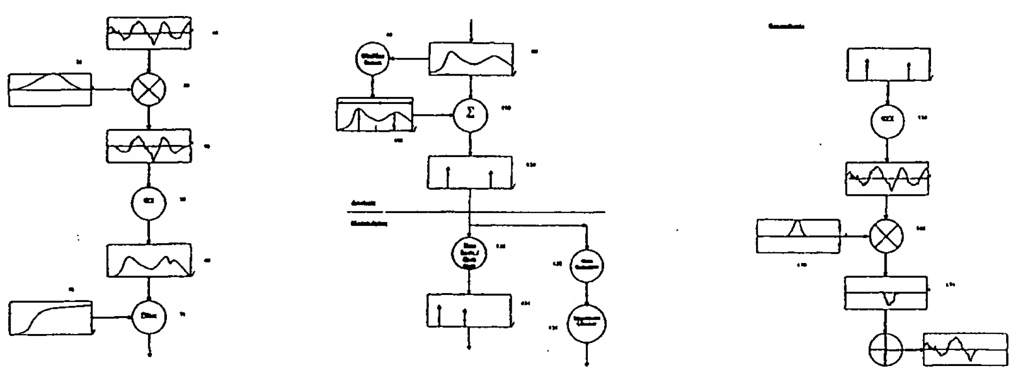
(72) Inventor; and
(75) Inventor/Applicant (*for US only*): HOEK, Stephen, Marcus, Jason [NZ/NZ]; 13 Vienna Place, Birkenhead, Auckland (NZ).

(74) Agents: HAWKINS, Michael, Howard et al.; Baldwin Shelston Waters, NCR Building, 342 Lambton Quay, Wellington (NZ).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), DM, EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SIGNAL PROCESSING TECHNIQUES FOR TIME-SCALE AND/OR PITCH MODIFICATION OF AUDIO SIGNALS



(57) Abstract

A method of signal processing for time scale and/or pitch modification of audio signals is disclosed. The method involves encoding and resynthesising a wave form whereby the wave form is sampled into a series of frames, each frame is multiplied by a windowing function where the peak of the windowing function is centred at approximately the zero point of each frame. The resulting function is then subjected to a Fast Fourier transform thus producing a frequency-domain wave form. The resultant wave form is convolved with a variable kernel function, the specification of the variable kernel function varying with frequency. Maxima and associated minima in a magnitude spectrum of each convolved frame are located so that each local maxima and associated minima define a plurality of regions. Each region corresponds to a frequency component of the signal. Each of the regions is analysed in the frequency domain representation separately by summing the complex frequency components or bins falling within the defined region to a signal vector. The variable kernel function can be usefully varied to achieve a differing trade of between the frequency and temporal resolution across the frequency range of the signal.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

SIGNAL PROCESSING TECHNIQUES FOR TIME-SCALE AND/OR PITCH MODIFICATION OF AUDIO SIGNALS

Field of the Invention

The present invention relates to encoding and manipulation of digital signals. More particularly, although not exclusively, the present invention relates to time-scale and/or pitch modification of audio signals. As such, the signal analysis and re-synthesis method described herein is not limited to audio signals. It is envisaged that the present invention may find application in the coding of other signals with the (wavelet-like) method described herein. An example of such an application includes image compression. Essentially the present invention may be applied where one wishes to simultaneously analyse different regions of the frequency domain with differing temporal/spatial resolutions

Background to the Invention

There are a number of existing techniques for time-scale/pitch modification of audio signals which are known in the art. These can be broadly classified as follows.

(a) Time domain methods:

These techniques attempt to estimate the fundamental period of a musical signal by detecting periodic activity in the audio signal. By this process, an input signal is delayed and multiplied by the undelayed signal, the product of which is then smoothed in a low pass filter to provide an approximate measure of the autocorrelation function. The autocorrelation function is then used to detect a nonperiodic signal or a weak periodic signal which might be hidden in

the noise. Once the fundamental period of the musical signal is found the process is repeated and the analysed sections of the signal are overlapped. A significant disadvantage in these techniques is that most audio signals do not have a fundamental period. For example polyphonic instruments, recordings with reverberation and percussion sounds do not have an identifiable fundamental period. Further, when applying such methods, transients in the music are repeated. This leads to notes having multiple starts and ends. Another problem with this technique is that overlapping of the delayed sections of the music can produce an audio effect that is metallic, mechanical or exhibits echo-like nature.

(b) Sinusoidal analysis methods:

These techniques assume that the input signal is made up from pure sinusoids. The inherent disadvantage of such a method is therefore self-evident.

Sinusoidal analysis techniques use Short Time Fast Fourier Transforms (FFT) to estimate the frequency of the component sinusoids. The derived signal is then synthesised with a bank of tone generators to produce the desired output. Short Time Fourier Analysis captures information about the frequency content of a signal within a time interval, governed by the Window Function chosen. A significant disadvantage of such techniques is that a single time-domain window is applied to all the frequency content of the signal, so the signal analysis cannot correspond accurately to human perception of the signal

content. Also, conventional sinusoidal analysis methods use a local maxima search of the magnitude spectrum to determine the frequency of the constituent sinusoids including consideration of relative phase changes between analysis frames. This technique ignores any side-band information located around each of the local maxima. The effect of this is to exclude any signal modulation occurring within a single analysis frame, resulting in a smearing of the sound and almost a complete loss of transients. An example of such a transient, in the audio context, is a guitar pluck.

(c) Phase vocoder methods:

This type of technique uses a Fast Fourier Transform as a large bank of filters and treats the output of each of the filters separately. The relative phase change between two consecutive analyses of the input is used to estimate the frequency of the signal content within each bin. A resulting frequency-domain signal is synthesised from this information, treating each bin as a separate signal. In contrast to sinusoidal analysis techniques, this method retains the spectral energy distribution of the original signal. However, it destroys the relative phase of any transient information. Therefore, the resulting sound is smeared and echo-like.

In view of the prior art techniques, it would therefore be desirable to analyse and process audio signals so that the resultant output retains the tonal characteristics of the original signal and is capable of accurately capturing

transient sounds without smearing or introducing an echo-like character to the output signal.

Accordingly, it is an object of the present invention to provide a technique for processing audio signals which achieves the above mentioned aims, ameliorates at least some of the disadvantages inherent in the prior art or at least provides the public with a useful choice. Further, it is an object of the invention to provide a signal analysis and synthesis method that can also be applied to the coding of signals in general.

Disclosure of the Invention

In one aspect the invention provides for a method of encoding and re-synthesising a waveform, the method including:

sampling the waveform to obtain a series of discrete samples and constructing therefrom a series of frames, each frame spanning a plurality of samples;

multiplying each frame with a windowing, preferably raised cosine, function wherein the peak of the windowing function is centred substantially at a zero point of each frame;

applying a Fast Fourier Transform to each frame thereby producing a frequency-domain waveform;

convoluting the resultant frequency domain data with a variable kernel function, whose specification varies with frequency;

locating local maxima and surrounding minima in the magnitude spectrum of each convolved frame, wherein each local maxima and associated minima define a plurality of regions, each region corresponding to a frequency component of the signal; and

analysing each of the regions in the frequency domain representation separately by summing the complex frequency components of bins falling within the defined region into a signal vector; wherein the variable kernel function can be usefully varied to achieve a differing tradeoff between frequency and temporal resolution across the frequency range of the signal.

In a preferred embodiment, the waveform corresponds to a digitised audio frequency waveform wherein the kernel function may be varied to approximate the perceptual characteristics of the human ear.

In the case where the waveform corresponds to an audio signal, the location of the maxima corresponds to the perceived pitch of the frequency component.

The method may further include the step of manipulating the signal while represented as signal vectors.

Such manipulation may take the form of modifying pitch or time scale (in an audio signal) or further data reduction adapted for efficient signal storage and/or transmission.

In the case of modifying an audio signal, the frequency location and phase of analysed signal vectors can be shifted as necessary to achieve a scaling of time and/or pitch.

Converting back to the sampled time domain representation of the signal may be achieved by accumulating into the frequency domain an equivalent signal whose components correspond to those signal vectors determined in the analysis of the original signal.

Preferably an Inverse Fast Fourier Transform may be applied so as to give a time domain signal that may be suitably windowed and accumulated to produce the decoded signal.

Preferably the form of the convolution function is determined empirically by subjectively assessing the quality of the synthesised output.

Preferably the application of the kernel function to the frequency domain data is implemented as a single-pole low-pass filter operation on said data, the pole's location being varied with frequency.

Preferably, in the case of the analysis of audio signals, the pole may be specified by a control function *s(f)* of the form:

$$s(f) = 0.4 + 0.26\arctan(4\ln(0.1f) - 18)$$

where f is the frequency in hertz (cycles per second).

The frequency domain filter may be specified by the relation:

$$y_{out}(f) = [1 - s(f)]y_{in}(f) + s(f)y_{out}(f-1)$$

Preferably, for the purposes of manipulating an audio signal, each signal vector is treated separately; for pitch shifting the frequency of the component is multiplied by a real-valued pitch factor; for both pitch shift and time scale modification the necessary phase shift for glitch free reconstruction is calculated and applied.

Preferably the method includes the further steps of:

zeroing a frequency domain output array, and for each analysed frequency component represented as an analysed signal vector;

mapping the real-valued frequency to the two nearest integer-valued frequency bins; and

distributing the analysed signal vector between the two bins in proportion to 1 minus the real-valued frequency and the respective bins' locations.

In an alternative aspect, the resulting regions may be translated in frequency, so that the location of the maxima is scaled while the surrounding region is translated.

For each region, having a maxima and a first and second associated minima, for pitch shifting of an audio signal, the location of each maxima in the frame is scaled by the pitch shift factor, and associated harmonic information between the first and second minima is translated to respective positions around the scaled maxima.

To time stretch or compress the signal, each maxima is retained in the same location in the frequency domain while the band of frequency domain or harmonic information associated with the maxima is stretched or compressed, thereby stretching the amplitude and frequency modulation of the harmonics while preserving the pitch of the input signal.

The method may further include the further steps of:

resampling the data in each of the frames into a plurality of bins;

mapping each bin to a real valued location in an output frame where for a bin x lying within a band with a maximum at a frequency $freq_{max}$ the real valued location in the output frequency domain is y, wherein

$$y = freq_{\max} \times shift + \frac{(x - freq_{\max})}{(scale)}$$

Where *shift* equals the frequency shift and *scale* equals time expansion ratio.

Preferably, y is rounded down to the nearest integer z which is less than or equal to y wherein output bins z and z+1 are then added to, in proportion to 1 minus the difference between y and that bins integer location.

In a further aspect, the invention provides for software adapted to perform the above-mentioned method.

In a further aspect, the invention provides for hardware adapted to perform the above-mentioned method.

Brief Description of the Drawings

The invention will now be described by way of example only and with reference to the drawings in which:

Figure 1: illustrates a simplified schematic block diagram of an embodiment of the method of the invention (split over pages 28 to 30);

Figure 2: illustrates a simplified schematic block diagram of an embodiment of the alternate method of the invention (split over pages 31 to 33);

Figure 3: illustrates a schematic diagram of the process of searching for the maxima/minima;

Figure 5a and 5b: illustrates pitch and time stretching in respect of two maxima.

Referring to figure 1, a simplified flowchart illustrates the overall steps in an embodiment of the method of signal processing. For clarity, the schematic is split over pages 15 to 17.

An input audio signal is digitised into frames 10. Each of these frames is then processed as follows:

Each frame 10 is windowed (20) with (for example) a wide cosine function 30 producing time domain modulated representation of the input signal frame 10. A Fast Fourier Transform 50 is then applied to the frame producing a frequency domain representation of the input signal 60.

The frequency domain data 60 is then filtered with a filtering function 71 parameterised by $s(f)$. The filtering function may also be viewed as a low-pass

$$y_{out}(f) = [1 - s(f)]y_{in}(f) + s(f)y_{out}(f-1)$$

single pole filter in the present example. The function $s(f)$ 70 specifies how the behaviour of the filter varies with frequency. The filtering function 71 can be described by the recursive relation:

Thus $s(f)$ controls the 'severity' of the filter 71. So in effect, a different convolution kernel is used for each frequency bin. The real and imaginary components of each bin are convolved separately. In the present exemplary embodiment, the filtering or convolution function 71 has the effect of "blurring" the frequency domain information and therefore the convolving function can be referred to as a blurring function. Blurring or spreading the frequency domain data corresponds to a narrowing of the equivalent window in the time domain frame. Therefore each frequency bin of the fast Fourier Transform is effectively calculated as if a different sized time domain window had been applied before the FFT operation.

The effect of the filter does not have to be to blur the data. For example, translating the time domain samples by half the window size would make it necessary to high-pass filter the frequency domain data, to achieve the same equivalent windowing in the time domain.

The frequency domain filter 71 is applied to each bin in ascending order and then applied in descending order of frequency bin. This is to ensure that no phase shift is introduced into the frequency domain data.

A key aspect of the present invention is that the control function *s(f)* is chosen, in the case of processing audio frequency data, so as to approximate the excitation response of human cilia located on the basilar membrane in the human ear. In effect, the function *s(f)* is chosen so as to approximate the time/frequency response of the human ear.

The form of the control function *s(f)* is, in the present preferred embodiment, determined empirically by gauging the quality of the output or synthesised waveform under varying circumstances. Although this is a subjective procedure, repeated and varied evaluations of the quality of the synthesised sound have been found to produce a highly satisfactory convolution function.

A preferred form of the control function *s(f)* is:

$$s(f) = 0.4 + 0.26\arctan(4\ln(0.1f) - 18)$$

where f is the frequency in hertz (cycles per second).

In effect, the aforementioned steps are analogous to an efficient way to process a signal through a large bank of filters where the bandwidth of each filter is individually controllable by the control function *s(f)*.

Once the filter 71 is applied, the convolved frequency domain data 80 is analysed (90) to determine the locations of local maxima and the associated local minima.

To perform this step, it has been found that it is more efficient to use the intensity spectrum. Therefore, for each frequency, the data is a local maximum if $I(f) > I(f-1)$ and $I(f) > I(f+1)$. Local minima exist if $I(f) < I(f-1)$ and $I(f) < I(f+1)$. Here, $Mag(f) = \sqrt{real(f)^2 + im(f)^2}$ and $Intensity(f) = real(f)^2 + im(f)^2$.

Referring to figure 2, each maxima and associated local minima is used to define regions (indicated by the shaded arrows in figure 3) which correspond to an audible harmonic in the original audio frequency signal. The location of the maxima in the frequency domain corresponds to the perceived pitch of the harmonic and the band of the frequency domain information around the maxima represents any associated amplitude or frequency modulations of that harmonic. Since it is important not to lose this information, a summation of the whole band of frequencies around the peak is used to give a signal vector. This way the temporal resolution of the analysis sample will match the bandwidth of any modulations taking place.

Each of the regions is processed separately accordingly to the following technique. An accurate estimate of the location of each maxima is determined. Referring to figure 3, lower graph, the large arrow *a* (300) is the difference between the smallest intensity of the three intensity arrows (max-1) and the maximum intensity (max). The small arrow *b* (310) is the difference between the smallest (max-1) and the intermediate intensity (max+1). The ratio of the two is used to offset the integer maximum value.

Pitch shifting and time-scale modification are indicated schematically in figure 1 by the numeral 130. At this point alternative applications are indicated by data reduction (133) or transmission/storage (134) steps. These are illustrated as alternative options in figure 1.

The manipulated data are re-synthesised according to the following method: For the ith analysed frequency component, vector(i) has a real-valued location y in the frequency domain output.

y is rounded down to the nearest integer which is less than or equal to y and denoted z. Thus $z = Int(y)$.

The output bins z and $z+1$ are then added to with vector(i), in proportion to 1 minus the difference between y and that bins integer location.

$$\mathrm{Bin}[z] = \mathrm{Bin}[z] + [1-(y-z)]\ \mathrm{vector}(i)$$

$$\mathrm{Bin}[z+1] = \mathrm{Bin}[z+1] + (y-z)\ \mathrm{vector}(i)$$

where all operations are carried out on complex numbers.

To modify the time-scale or pitch of the analysed signal, it is necessary to compensate for any phase shifts so that the synthesised output is consistent (i.e. glitch free). To this end, the output signal in any one frame is moved forward in time by a fixed number of samples. Therefore, for a given pitch measurement it is possible to determine how much the output phase should change so that that the output smoothly joins with the previously synthesised frame.

However, the input time frame is moving by some other number of samples. Therefore, the analysed phase values are already changing as the analysis window moves through the input data.

Therefore the difference between the rate of change of input phase and the required rate of change of output phase is calculated. The difference between these phases is a measure of how fast to rotate the phase of the frequency domain data between analysis and synthesis. Each of the signal vectors defined above has a frequency measurement. This measurement is used to calculate how quickly to spin a vector of magnitude 1, where the vector is a complex number of representation. This vector is multiplied by the signal vector to provide the necessary phase shift for synthesis without affecting the timing of the decay characteristics or other modulations for each region.

This phase shift (in radians) is given by:

$$phase(i) = \frac{(2\pi f[t_r - t_a])}{t_w}$$

Where t_r = reconstruction time step in samples, t_a = analysis time step in samples and t_w = FFT size in samples.

Since the measurement of frequency provides a measure of phase difference between one synthesis frame and the next, these differences must be summed cumulatively as synthesis proceeds.

The cumulative sum applies only to one region, therefore regions must be tracked from one synthesis frame to the next.

A convenient data structure has been developed to track regions from one frame to the next and is described with reference to figure 4a and 4b. One integer array contains the location of the local maximum within a region for all the bins in that region. A corresponding array contains the last phase value (in radians) used to rotate that regions phase. The phase value is stored in the bin with the same index as the location of the maximum.

Therefore, when a new frame is analysed and local maxima detected, the location of the maximum is used to index into the integer array. This provides

the index of the maximum that existed in the previous frame. This index is then used to access the array containing the last phase value used for the corresponding region in the previous synthesis frame. This is illustrated in figures 3a and b whereby an analysis frame n is illustrated along with the nearest maxima array and the phase array. Considering the n+1 analysis frame, the first frequency maxima is 7. The corresponding seventh element of the nearest maxima array from the previous frame is 5. The fifth element of the phase array frame from the previous frame n is 12 degrees. This is updated using an estimate of the local maxima and then stored in the phase array for the next frame using position 7. For the second region 410 the thirteenth element of the nearest maxima array from the previous analysis frame n gives 16. From the phase array of the previous analysis frame n the phase is given as 57 degrees. A frequency estimate is used to update this phase value and is placed in the position 13 of the next phase array.

A frequency domain representation of the signal is constructed from the known signal components. For each signal vector, that vector is added to the frequency domain output array. Since the frequency locations are real valued, the energy from a signal vector is distributed between the nearest two (integer valued) bin locations. The frequency domain representation is then inverse Fourier transformed (150 in figure 1 page 16) to provide a time domain representation of the synthesised signal. Since the signal was analysed with differing temporal resolutions at different frequencies, the synthesised time

domain signal is only valid in the region equivalent to the highest temporal analysis resolution used. To this end, the synthesised time domain signal is windowed (160) with a (relatively) small positive cosine window (170), before being added (172) in an overlapping fashion to the final synthesised signal (180).

A variation, although equivalent, method of manipulating the information to achieve pitch shifting and time stretching is as follows.

The alternate method is substantially similar to the first method, sharing identically the steps of windowing (420), Fourier transforming (450), filtering (460), minima and maxima detection (490). The major difference between the two methods is after this point. Whereas the first method sums the contents of each region into a signal vector (110), the alternate method instead explicitly retains the contents of each region (510). The contents of each region are then translated and scaled in accordance with the pitch shift and time stretch factors respectively (530). For a pitch shift operation, the contents of a region are translated such that the maximum is scaled in frequency. For a time stretch operation, the contents of a region are scaled by the time stetch factor, but so that the maximum does not change in frequency.

Phase shift compensation is carried out substantially as described above with reference to figure 4a and 4b. To synthesise the output, the frequency domain data to be synthesised is copied a region at a time from the unaltered output of

Fourier transform step. The contents of each region are accumulated into the output frequency domain buffer in the same fashion as the first method.

There exist variations in the implementation of these two techniques that will be clear to one skilled in the art. However, the key feature of the present invention resides in using a control function $s(f)$ to vary a frequency domain filter at different frequencies. This brings about a windowing effect on the equivalent time-domain data that varies with frequency. In the case of processing audio frequency signals, this control function is chosen to reflect the response of the human cilia to a range of audio frequencies. Although the shape of this curve is determined empirically, it is possible that other curves may prove suitable for other manipulative techniques and applications.

A further feature of the present invention resides in the identification and location of the maxima and associated minima. The presently disclosed technique is computationally highly efficient and allows rapid high quality time stretching and pitch shifting of audio signals.

Experimentally, it has been shown that the present technique produces a sound with significantly enhanced tonal qualities and it is believed that this is largely achieved through the preservation of the harmonic information in the side-bands of the local frequency maxima.

In terms of a practical implementation of the present invention, it is envisaged that the technique may be implemented in software or alternatively in hardware. In the latter case, the hardware may form part of an audio component such as an audio player. Potential applications of the invention include the sound recording industry where audio signal processing/synthesis is commonly required to meet very high standards of reproduction quality. Alternative applications include those in the entertainment industry and it is anticipated that the technique may find application in sound reproduction/transmission systems where variations in pitch or tempo may be desirable. It is further anticipated that applications may exist in general signal processing, data reduction and/or data transmission and storage. In the latter case, the selection of the particular convolution function may vary.

Where in the foregoing description reference has been made to elements or integers having known equivalents, then such equivalents are included as if they were individually set forth.

Although the invention has been described by way of example and with reference to particular embodiments, it is to be understood that modifications and/or improvements may be made without departing from the scope of the appended claims.

CLAIMS:

1. A method of encoding and re-synthesising a waveform, the method including the steps of:

sampling the waveform to obtain a series of discrete samples and constructing therefrom a series of frames, each frame spanning a plurality of samples;

multiplying each frame with a windowing function wherein the peak of the windowing function is centred substantially at a zero point of each frame;

applying a Fast Fourier Transform to each frame thereby producing a frequency-domain waveform;

convoluting the resultant frequency domain data with a variable kernel function, the specification of the variable kernel function varying with frequency;

locating local maxima and surrounding minima in the magnitude spectrum of each convolved frame, wherein each local maxima and associated minima define a plurality of regions, each region corresponding to a frequency component of the signal; and

analysing each of the regions in the frequency domain representation separately by summing the complex frequency components or bins falling within the defined

region into a signal vector; wherein the variable kernel function can be usefully varied to achieve a differing tradeoff between frequency and temporal resolution across the frequency range of the signal.

2. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the windowing function is a raised cosine function.

3. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the waveform corresponds to a digitised audio frequency waveform wherein the kernel function is varied to approximate the perceptual characteristics of the human ear.

4. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the waveform corresponds to an audio signal, and the location of the maxima corresponds to the perceived pitch of the frequency component.

5. A method of encoding and re-synthesising a waveform as claimed in claim 1 further including the step of manipulating the signal while represented as signal vectors.

6. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein said manipulation takes the form of modifying pitch or time scale (in an audio signal) or further data reduction adapted for efficient signal storage and/or transmission.

7. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein, in the case of modifying an audio signal, the frequency location and phase of analysed signal vectors are shifted according to a predetermined amount to achieve a scaling of time and/or pitch.

8. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein converting back to the sampled time domain representation of the signal is achieved by accumulating into the frequency domain an equivalent signal whose components correspond to those signal vectors determined in the analysis of the original signal.

9. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein an Inverse Fast Fourier Transform is

applied so as to give a time domain signal that may be suitably windowed and accumulated to produce the decoded signal.

10. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the form of the convolution function is determined empirically by subjectively assessing the quality of the synthesised output.

11. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the application of the kernel function to the frequency domain data is implemented as a single-pole low-pass filter operation on said data, the pole's location being varied with frequency.

12 A method of encoding and re-synthesising a waveform as claimed in claim 11 wherein, in the case of the analysis of audio signals, the pole is specified by a control function $s(f)$ of the form:

$$s(f) = 0.4 + 0.26\arctan(4\ln(0.1f) - 18)$$

where f is the frequency in hertz (cycles per second).

13. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the frequency domain filter may be specified by the relation :

$$y_{out}(f) = [1 - s(f)]y_{in}(f) + s(f)y_{out}(f-1)$$

14. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein, for the purposes of manipulating an audio signal, each signal vector is treated separately; for pitch shifting the frequency of the component is multiplied by a real-valued pitch factor; for both pitch shift and time scale modification the necessary phase shift for glitch free reconstruction is calculated and applied.

15. A method of encoding and re-synthesising a waveform as claimed in claim 1 wherein the method includes the further steps of:

zeroing a frequency domain output array, and for each analysed frequency component represented as an analysed signal vector;

mapping the real-valued frequency to the two nearest integer-valued frequency bins; and

distributing the analysed signal vector between the two bins in proportion to 1 minus the real-valued frequency and the respective bins' locations.

16. A method of encoding and resynthesizing a waveform as claimed in claim 1 wherein the resulting regions in the frequency domain are translated around each maxima to a different frequency, the position of the maxima and the resulting signal being a multiple of the frequency of the maxima so that the location of the maxima is scaled while the surrounding region is translated.

17. A method of encoding and resynthesizing a waveform as claimed in claim 16 wherein for each region, having a maxima and a first and second associated minima, for pitch shifting of an audio signal, the location of each maxima in the frame is scaled and associated harmonic information between the first and second minima and maxima is translated to respective positions around the maxima.

18. A method of encoding and resynthesizing a waveform as claimed in claim 16 or 17 wherein to time stretch the signal, each maxima is retained in the same location in the frequency

domain while the band of frequency domain or harmonic information associated with the maxima is compressed, thereby stretching the amplitude and frequency modulation of the harmonics while preserving the pitch of the input signal.

19. A method of encoding and resynthesizing a waveform as claimed in claim including the further steps of:

resampling the data in each of the frames into a plurality of bins;

mapping each bin to a real valued location in an output frame where for a bin x lying within a band with a maximum at a frequency $freq_{max}$ the real valued location in the output frequency domain is y, wherein

$$y = freq_{\max} \times shift + \frac{(x - freq_{\max})}{(scale)}$$

where *shift* equals the frequency shift and *scale* equals time expansion ratio.

20. A method of encoding and resynthesizing a waveform as claimed in claim 19 wherein, y is rounded down to the nearest integer z which is less than or equal to y wherein output bins z and $z+1$ are then added to, in proportion to 1 minus the difference between y and that bins integer location.

21. A software application operating in accordance with the method of claims 1 to 20.

22. A device constructed to perform in accordance with the method of claims 1 to 20.

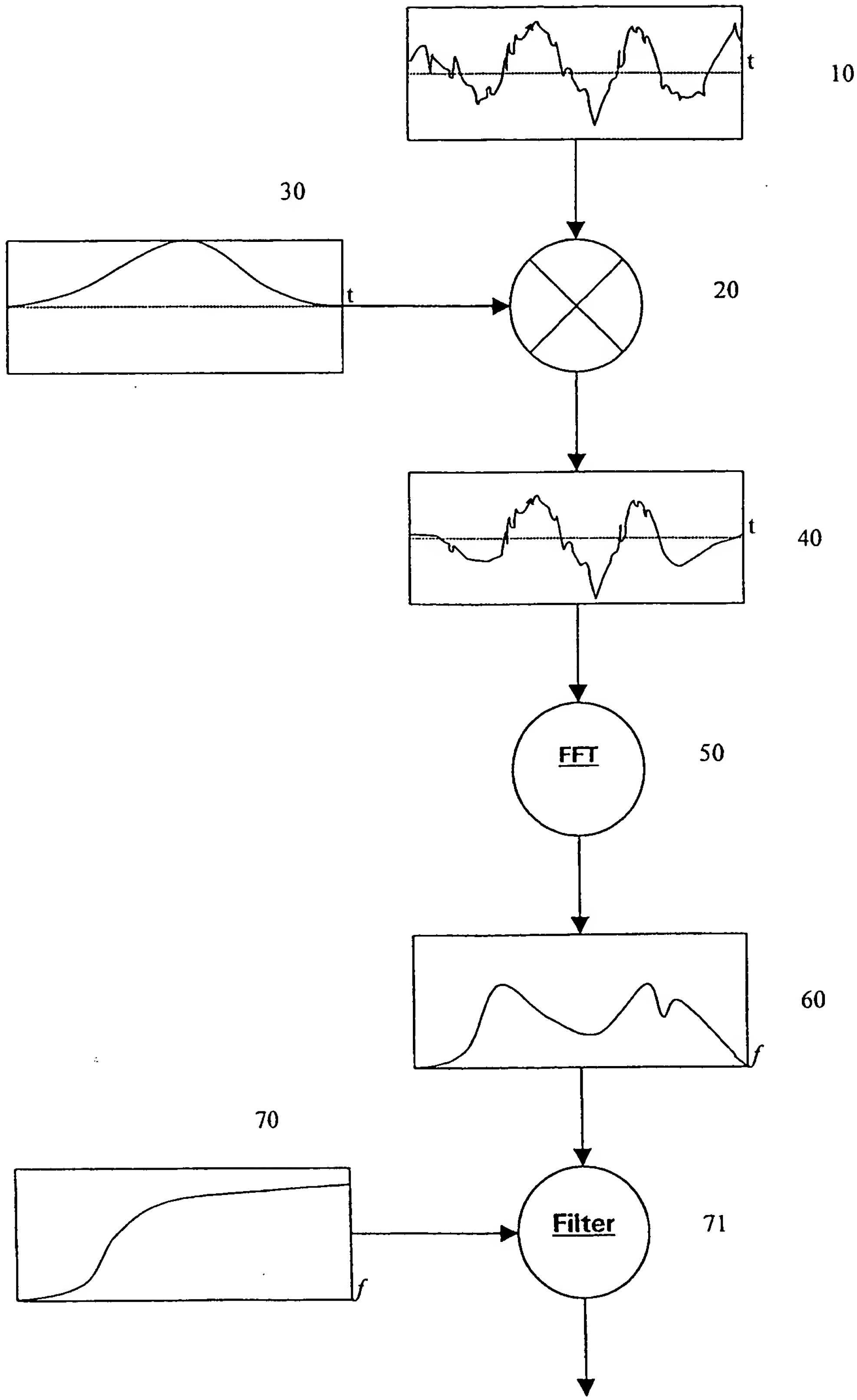


FIG 1

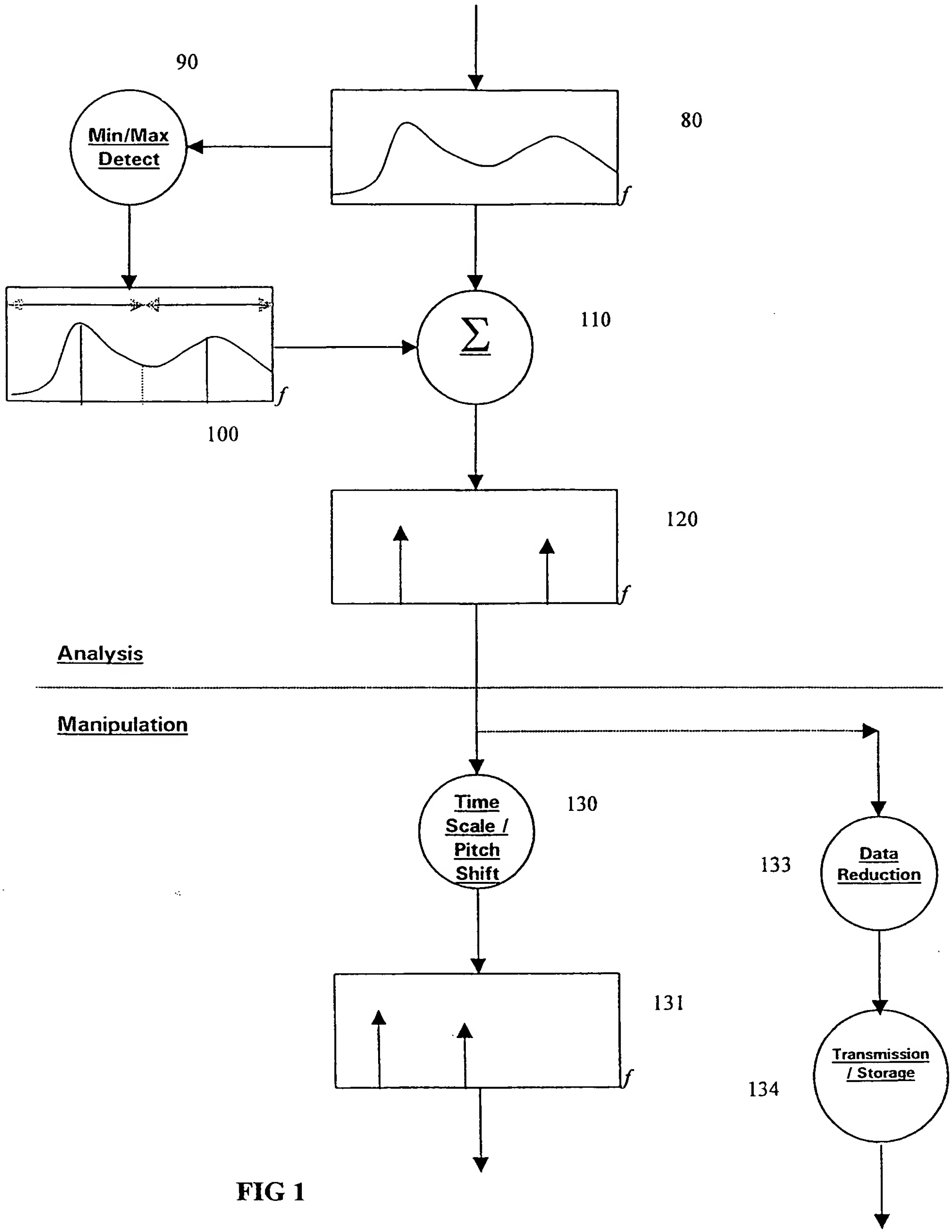

90
Min/Max Detect
80
f
100
f
Σ
110
120
f
Analysis
Manipulation
Time Scale / Pitch Shift
130
133
Data Reduction
131
f
134
Transmission / Storage

FIG 1

Resynthesis

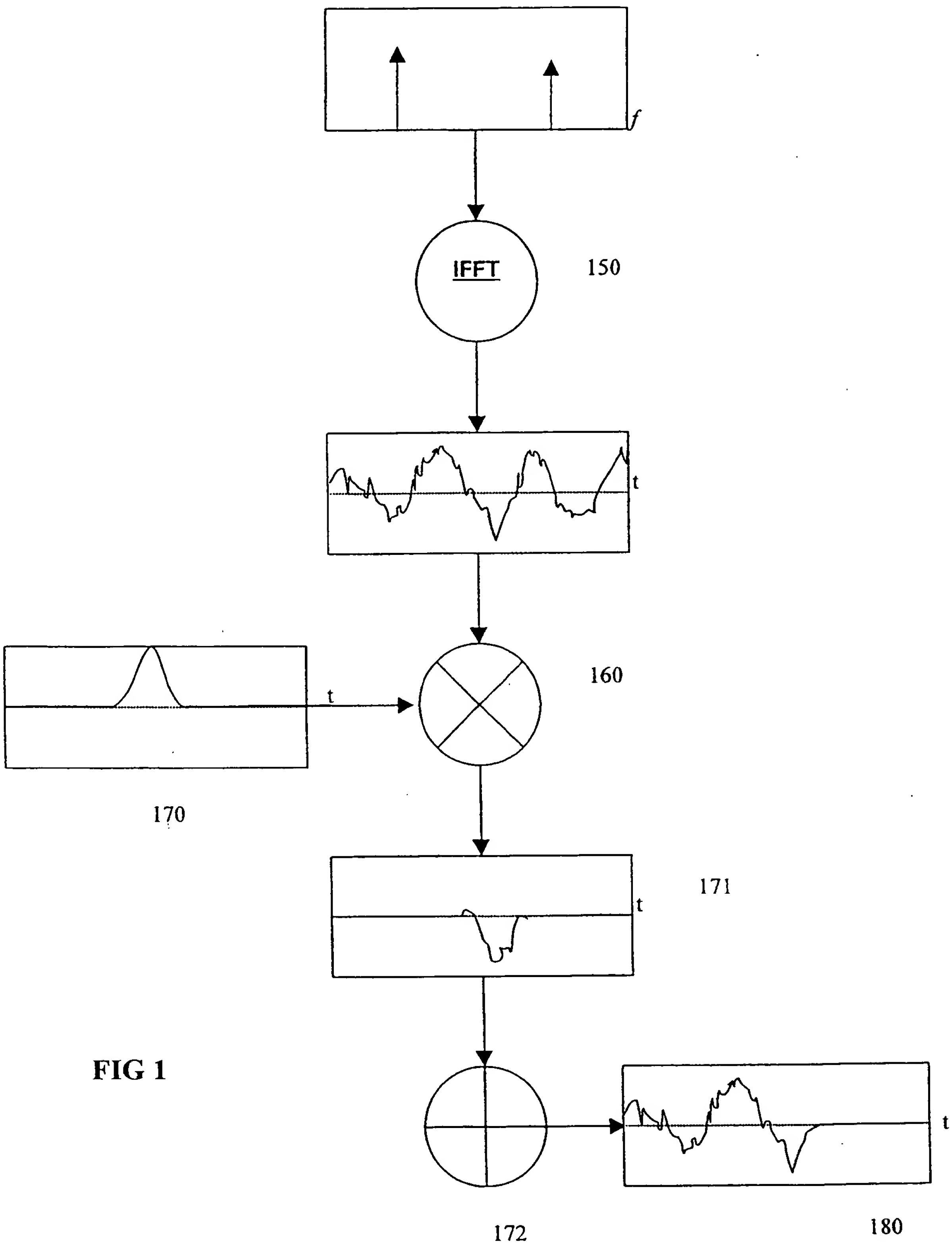


FIG 1

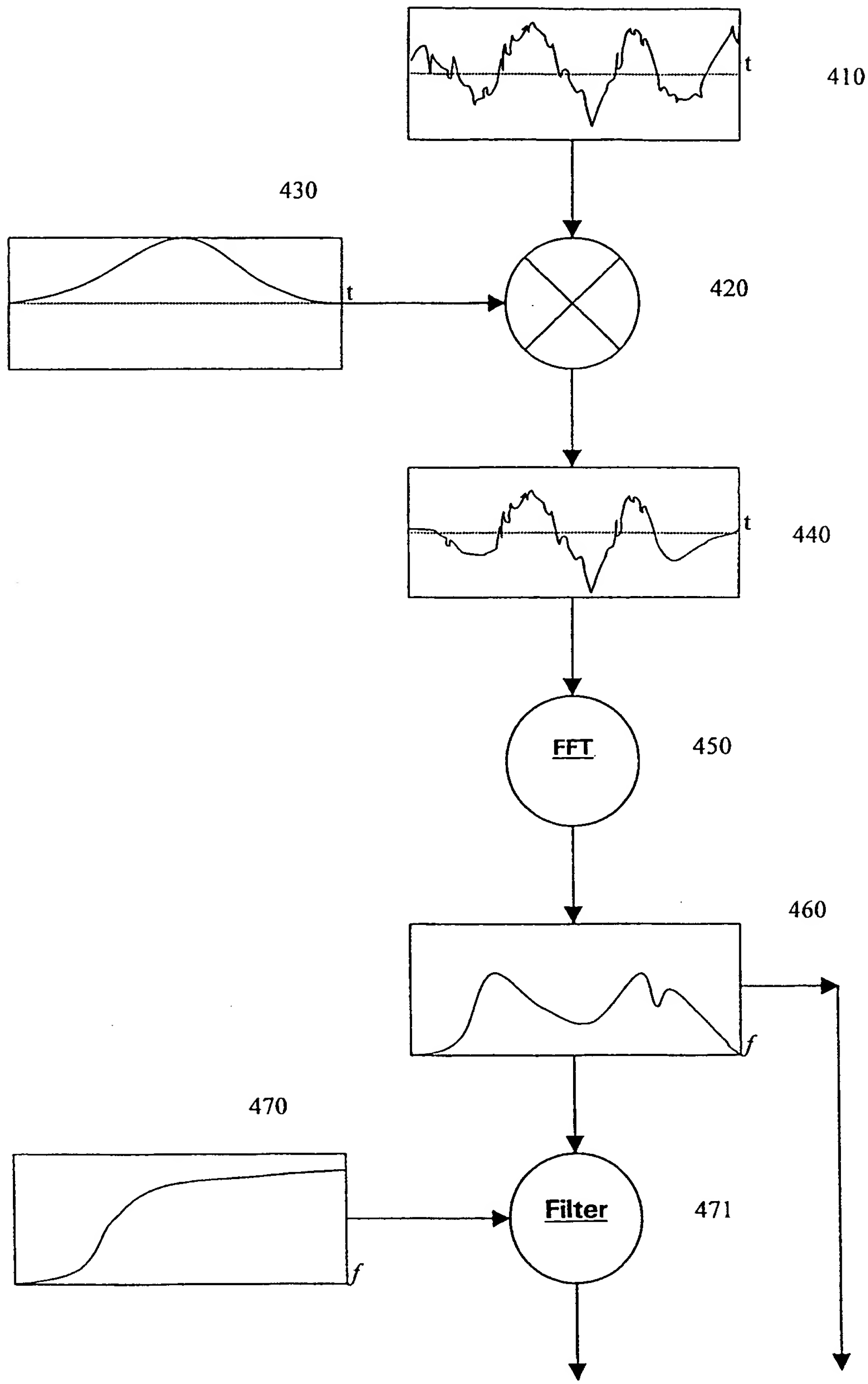


FIG 2

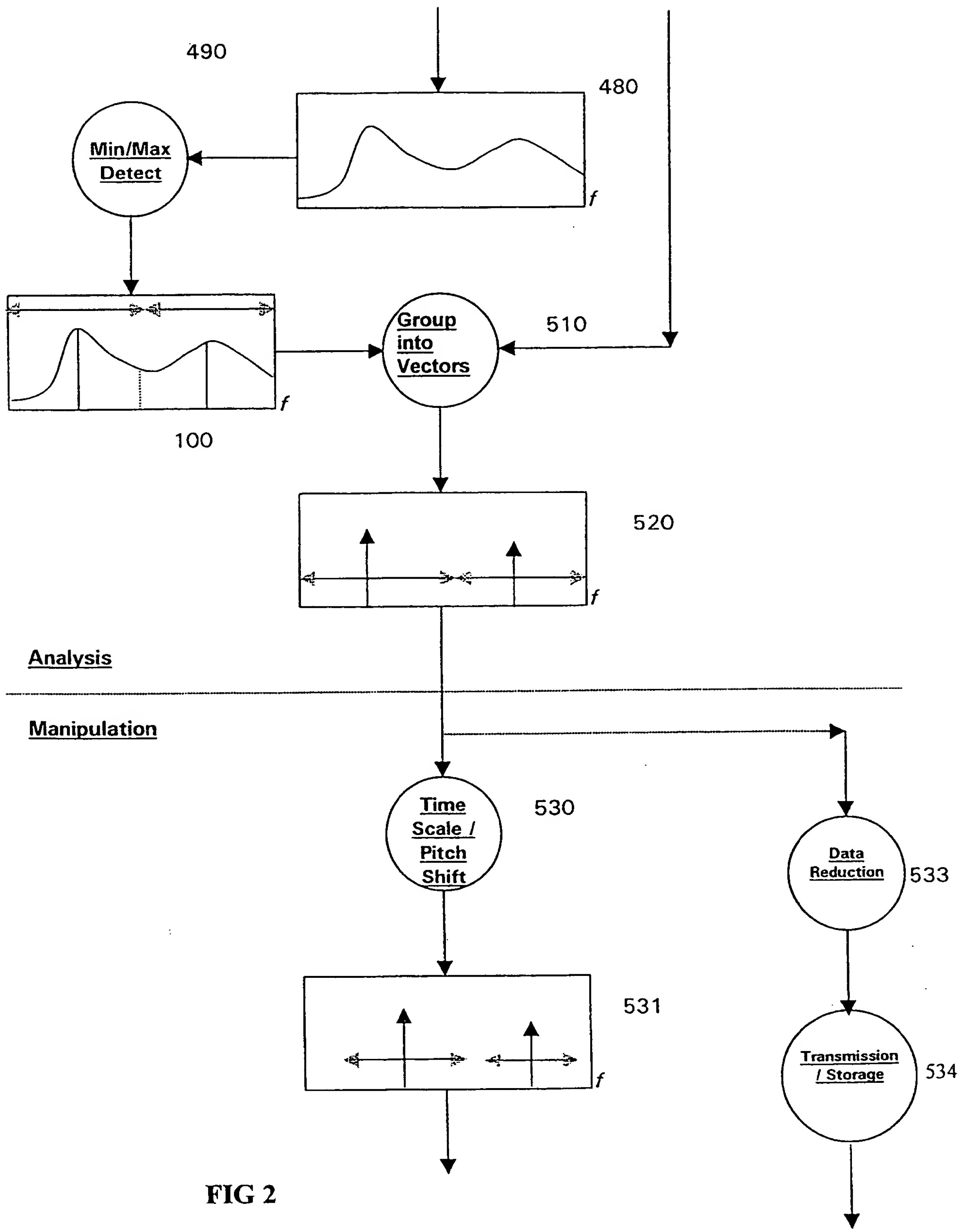

490
480
Min/Max
Detect
f
Group
into
Vectors
510
f
100
520
f
Analysis
Manipulation
Time
Scale /
Pitch
Shift
530
Data
Reduction
533
531
f
Transmission
/ Storage
534

FIG 2

Resynthesis

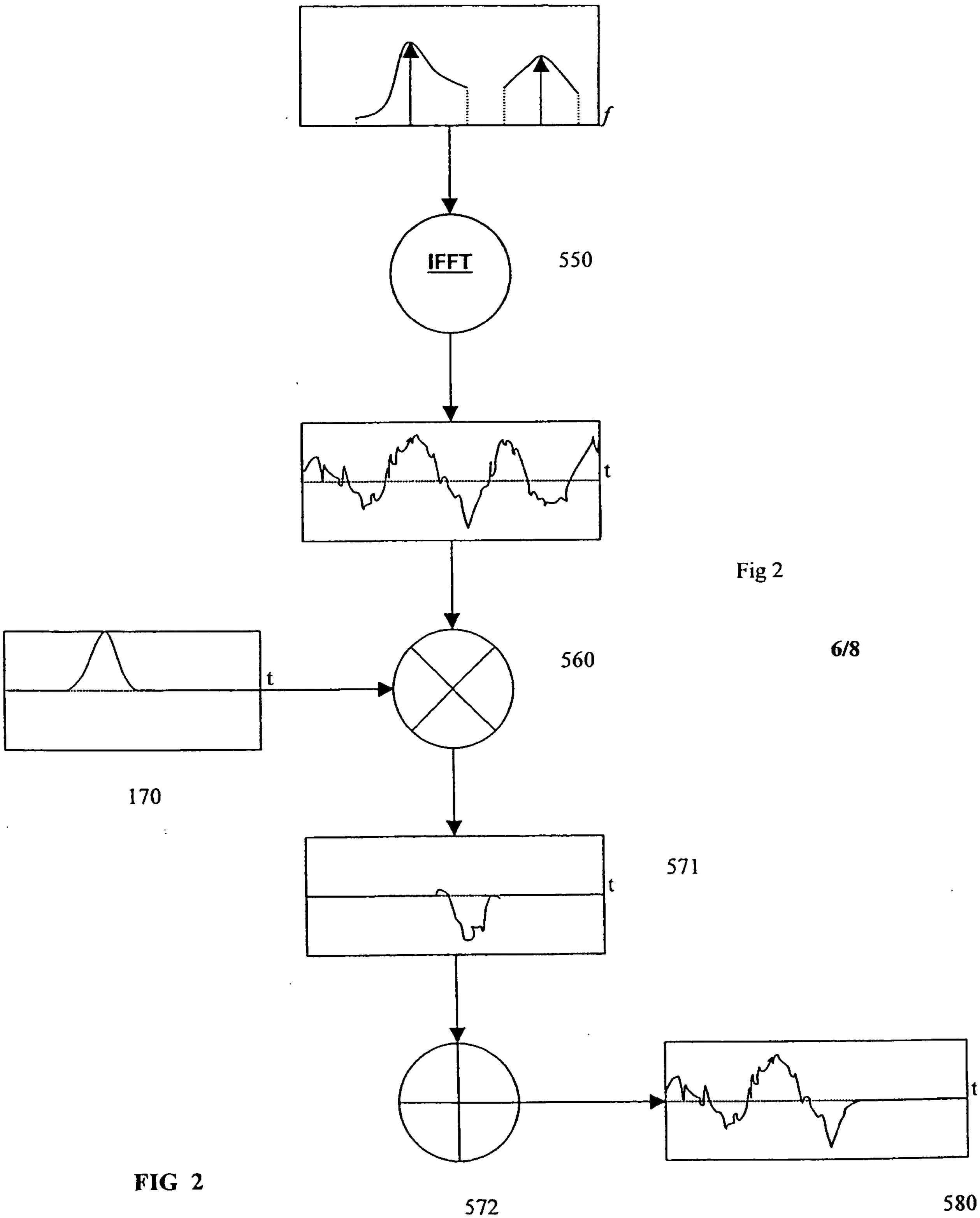


Fig 2

6/8

FIG 2

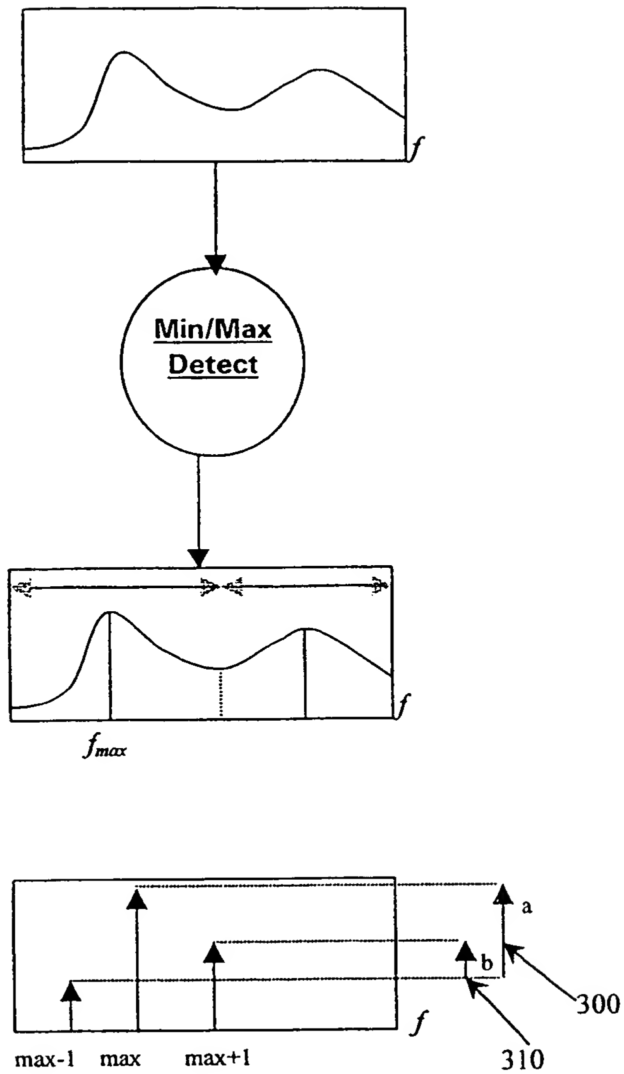
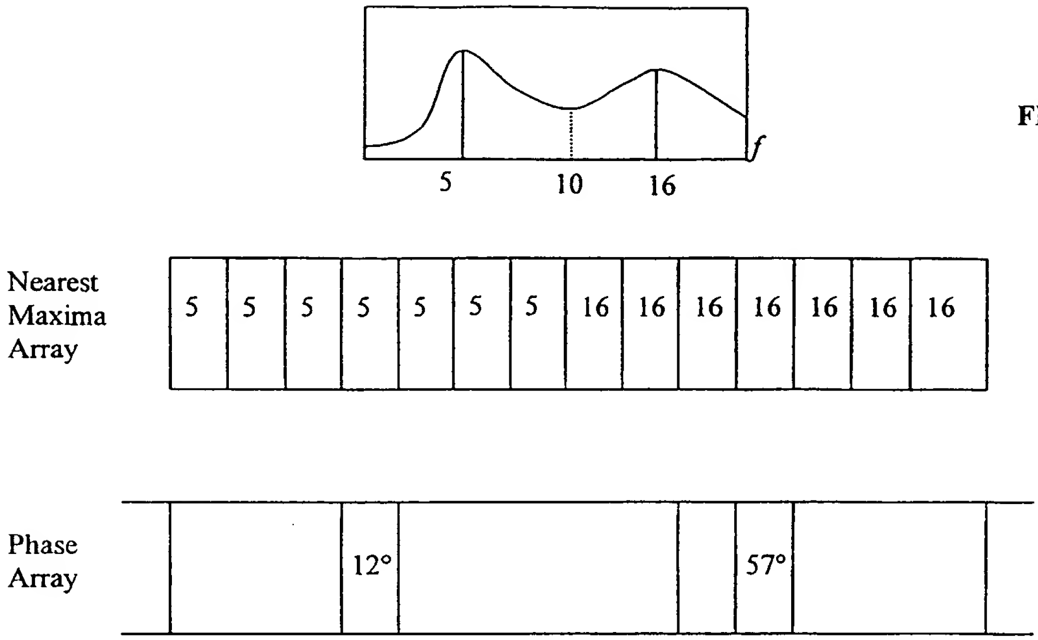
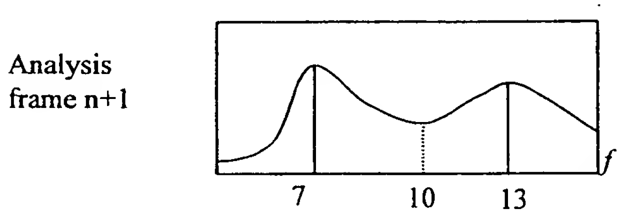


FIG 3



f
5
10
16
Nearest Maxima Array
5 5 5 5 5 5 5 16 16 16 16 16 16 16
Phase Array
12°
57°

FIG 4(a)



Analysis frame n+1
f
7
10
13

FIG 4(b)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/NZ 99/00143

A.	CLASSIFICATION OF SUBJECT MATTER
Int Cl[6]:	G10L 7/06, 9/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC H04N, G10L, G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 250048 A (N.V. Philips) 23 December 1987 Whole document	1-22

☐ Further documents are listed in the continuation of Box C

☒ See patent family annex

* Special categories of cited documents:

"A" Document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
23 December 1999	10 JAN 2000

Name and mailing address of the ISA/AU
AUSTRALIAN PATENT OFFICE
PO BOX 200
WODEN ACT 2606 AUSTRALIA
E-mail address: pct@ipaustralia.gov.au
Facsimile No.: (02) 6285 3929

Authorized officer
J. LAW
Telephone No.: (02) 6283 2179

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/NZ 99/00143

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
EP	250048	AU	74466/87	CA	1266893	JP	63004710
		NL	8601604	US	4807173		

END OF ANNEX